Twitter Thread by Vladimir Haltakov



Vladimir Haltakov
@haltakov

This is Karma. Karma is not a machine learning classifier ■■■

Karma is a real dog trained to detect drugs. However, he would fail the simplest tests we apply in ML...

Let me take you through this story from the eyes of an ML engineer.

https://t.co/WAXRUlTvSI

Thread ■

Story TLDR ■

The story is about police dogs trained to sniff drugs. The problem is that the dogs often signal drugs even if there are none. Then innocent people land in jail for days.

The cops even joke about the “probable cause on four legs”.

Let's see why is that ■

1. Sampling Bias ■

Drugs were found in 64% of the cars Karma identified, which was praised by the police as very good. In the end, most people don't carry drugs in their cars, so 64% seems solid.

There was a sampling problem though... ■

The cars were not sampled at random! The police only did the sniff test if there was a serious suspicion that something is wrong.

The chance there are drugs in the car is much higher in this case!

2. Evaluation Metrics ■

The police referred to a 2014 study from Poland measuring the efficacy of sniffer dogs. The problem was that every test actually contained drugs!

This means there was no chance to measure false positives from the dogs! Only recall, not precision ■■■■

3. Leaking Training Data ■

Another study found that the dogs learned to recognize the emotions of their handlers during tests. They felt that their human wanted them to find drugs in the specific test scenario, so they did.

The trainer leaked the ground truth during testing.

4. Overfitting ■

Similar to the one above, in many cases, the dog saw that their handler wanted to find drugs in a car during a traffic stop. So it would raise an alarm.

The dog was rewarded before the car was actually searched! It found an easy signal giving it a reward.

Summary ■

It is fascinating how many problems there are with the sniffer dogs that are well known to machine learning engineers (and of course mathematicians). Some of them are even common sense...

Avoid these problems not only when training your model, but also in life ■

If you liked this thread and want to read more about self-driving cars and machine learning follow me @haltakov!